Problem Set 4

REPLACE HERE WITH YOUR NAME

2024-11-23

Table of Contents

## Context

This problem set is inspired by the recent research article *“Women’s Empowerment and Child Mortality”* by Kellard et al. (2024), published in *World Development*.

## Question 1: Understanding the Dataset (2 points)

### 1a. Unit of Observation

What is the unit of observation in this dataset?

**Answer:**

### 1b. Variables

Describe each variable’s type (use the classification discussed in Unit I).

**Answer:**

### 1c. Number of Observations

How many observations are present in the dataset?

**Answer:**

# Write code to check the number of observations  
str(df)

## 'data.frame': 168 obs. of 11 variables:  
## $ female\_legislators\_percent : num 37.4 11.8 43.5 30.5 10.9 ...  
## $ avg\_female\_HoS\_or\_HoG : num 0 0 0 0.6 0 0 0 0 0 0 ...  
## $ iso3c : chr "MEX" "SUR" "SWE" "CHE" ...  
## $ country\_name : chr "Mexico" "Suriname" "Sweden" "Switzerland" ...  
## $ political\_geographic\_region : int 2 10 5 5 4 4 6 7 1 1 ...  
## $ year : int 2014 2014 2014 2014 2014 2014 2014 2014 2014 2014 ...  
## $ adolescent\_fert\_rate\_mean2015\_2018 : num 64.33 59.85 4.42 2.67 67.79 ...  
## $ infant\_mortality\_rate\_mean2015\_2018 : num 13.32 17.55 2.27 3.65 37.3 ...  
## $ life\_expect\_birth\_mean2015\_2018 : num 74.3 71.8 82.4 83.5 63.8 ...  
## $ one\_or\_more\_female\_HoS\_or\_HoG\_1994to2014: int 0 0 0 1 0 0 0 0 0 0 ...  
## $ region\_category\_factor : chr "Latin America" "The Caribbean" "Western Europe and North America" "Western Europe and North America" ...

### 1d. Summary Statistics

For all continuous and binary variables, create a table showing their mean, minimum, maximum, and standard deviation. Describe your findings.

# Use functions such as dyplr's `summarise()` or psych's `describe()` to compute summary statistics for the continuous and binary variables

**Answer:**

## Question 2: Hypothesis Testing (5 points)

In this question, we aim to test whether countries with at least one female head of state or government between 1994 and 2014 have lower infant mortality rates.

### 2a. Relevant Variables

Identify the variables necessary to test this hypothesis. Explain your answer.

**Answer:**

### 2b. Hypothesis Formulation

Write the null hypothesis () and the alternative hypothesis () in mathematical form. Clearly define any parameters or terms used in the hypotheses. Indicate whether this is a one-sided or two-sided test. Explain your answer.

**Answer:**

### 2c. Test Statistic

Assume we cannot rely on the central limit theorem due to a limited sample size. In this case, which test statistic should we use to test the null hypothesis? Explain your reasoning and write the formula for the test statistic. Explain each term of the formula and its interpretation.

**Answer:**

### 2d. Compute the Test Statistic

Using the dplyr function filter(), and the R functions mean(), and nrow(), compute the value of the test statistic.

# Write code to filter data and compute the necessary values for the test statistic

**Answer:**

### 2e. Compute the P-Value and Reach a Conclusion

* Calculate the p-value of the test statistic. Explain what the p-value is capturing and how we should interpret it.
* At the 5% significance level, can we reject the null hypothesis? Interpret and explain the conclusion of the hypothesis test.

# Use appropriate R function to compute the p-value (hint: see PS4's tutorial)

**Answer:**

## Question 3: Linear Regression (4 points)

This question examines the relationship between the 2015–2018 average infant mortality rate (dependent variable) and the 2014 percentage of female legislators.

### 3a. Hypothesis Formulation

Suppose we expect a positive relationship between the two variables. Write the hypothesis test for the population slope (). Indicate whether this is a one-sided or two-sided test.

**Answer:**

### 3b. Regression Estimation

Using the command lm(...), estimate the simple linear regression model. Interpret the estimated slope coefficient. How does the dependent variable change with a one percentage point increase in female legislators?

# Estimate the linear model (Hint: see the tutorial on Brightspace)

**Answer:**

### 3c. Test Statistic and P-Value

From the regression output, identify the test statistic and p-value for the estimated slope. Explain how to interpret these values.

**Answer:**

### 3d. Statistical Significance

At the 5% significance level, can we reject the null hypothesis? Interpret and explain the conclusion of the hypothesis test.

**Answer:**